A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation

Year: 2023 | Citations: 340 | Authors: A. Khan, Omkar Chaudhari, Rohitash Chandra

---

Abstract

Class imbalance (CI) in classification problems arises when the number of observations belonging to one class is lower than the other. Ensemble learning combines multiple models to obtain a robust model and has been prominently used with data augmentation methods to address class imbalance problems. In the last decade, a number of strategies have been added to enhance ensemble learning and data augmentation methods, along with new methods such as generative adversarial networks (GANs). A combination of these has been applied in many studies, and the evaluation of different combinations would enable a better understanding and guidance for different application domains. In this paper, we present a computational study to evaluate data augmentation and ensemble learning methods used to address prominent benchmark CI problems. We present a general framework that evaluates 9 data augmentation and 9 ensemble learning methods for CI problems. Our objective is to identify the most effective combination for improving classification performance on imbalanced datasets. The results indicate that combinations of data augmentation methods with ensemble learning can significantly improve classification performance on imbalanced datasets. We find that traditional data augmentation methods such as the synthetic minority oversampling technique (SMOTE) and random oversampling (ROS) are not only better in performance for selected CI problems, but also computationally less expensive than GANs. Our study is vital for the development of novel models for handling imbalanced datasets.